DOCUMENT RESUME

ED 183 622 TM 800 044

AUTHOR Halpin, Glennelle: And Others
TITLE Retention in an Actual Classroom Setting as a Function of Type and Complexity of Tests.
PUB DATE [Apr 79]
NOTE 13p.: Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, CA, April 8-12, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement: *Complexity Level: Higher Education: Knowledge Level: *Multiple Choice Tests: Performance Factors: *Retention: *Scores: Test Items
IDENTIFIERS *Short Answer Tests: *Test Format

ABSTRACT

Forty-five graduate students in two educational psychology classes were randomly assigned to three groups for a classroom test on the course content. One group was asked to go to another room and then was privately excused from the test. The other two groups took tests with the same 32 items, but one group used a multiple-choice format and the other group a short answer format. Four weeks later, all three groups were given both forms of the same test (the short answer form first) on a unannounced basis. In this test one half the items were measures of knowledge and one half were measures of concepts. Resulting scores were analyzed using a 3x2x2 factorial analysis of variance design with repeated measures in the last two factors. The treatment effect was not significant. Scores were higher on the multiple-choice test than on the short-answer test. There also was a significant interaction between test format and question complexity (knowledge questions versus concept questions); that is, concept questions were comparatively easier in multiple-choice form and harder in completion form. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED183622

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Retention in an Actual Classroom Setting as a Function of Type and Complexity of Tests

Glennelle Halpin and Gerald Halpin

Auburn University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Glennelle Halpin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Joseph Harrington

Troy State University

Paper presented at the meeting of the American Educational Research Association, San Francisco, April 1979.

TM800 044

Retention in an Actual Classroom Setting
as a Function of Type and Complexity of Tests

In 1885 Ebbinghaus published the first experimental study of human memory. There have been numerous studies in the fields of education and psychology since that time which have addressed the problems of learning and retention. These studies generally fall into two basic categories: (a) research concerned with theory which focused on the recall of word pairs, series of words, sentences, and the like and (b) research that took a more practical approach which was related to educational achievement and the retention of subject matter. It is studies in the latter category that are most relevant here.

A survey of the literature revealed a definite lack in studies of specific classroom practices related to retention of student learning. Harnqvist (1977) highlighted this deficiency in an invited address to the 1977 meeting of the American Educational Research Association which he called "Enduring Effects of School--A Neglected Area in Educational Research."

Our study is one in this neglected area. It was designed to investigate the effects of a particular classroom activity--that of merely taking a test--on student retention. Others have studied the effects of questioning on retention. Peeck (1970), for example, studied the effects of prequestions on delayed retention of prose. Ryan (1973) and Gall, Ward, Berliner, Cahen, Winne, Elashoff, and Stanton (1978) studied the effects of verbal questioning by "teachers" during "instruction." Roderick and Anderson (1968) and Anderson, Surber, Biddle, Zych, and Lieberman (Note 1) studied retention of information as a function of the nature, timing, and number of quizzes. None of these studies were conducted in actual classroom settings with both instructor and students going about the day by day business of teaching and learning. Also, in none of these studies were the effects of recall versus recognition questions at the knowledge versus concept levels investigated.

Specific objectives of this study were to determine if three groups of subjects taking a recall test, a selection test, or no test differed in test performance after a 4-week period and whether the treatment effects were moderated by the complexity of questions (knowledge vs. concepts) and the type of questions (recall vs. selection) utilized on the criterion measure.

Method

Subjects

Graduate students (N = 45) from two advanced educational psychology classes participated in this study. The classes were part of the summer program at a large southern university which attracts students with diverse backgrounds from an extensive geographical area. Both master's and doctoral students from a wide range of academic disciplines were represented.

Classroom Conditions

Learning theory was routinely taught by the same white female instructor using a traditional lecture-discussion-textbook approach. At the beginning of the course, an outline was distributed depicting the topics to be covered day by day along with the appropriate textbook reading assignments and a notation of the date on which the summative test on learning theory would be administered. On this date all students reported to take the test.

Treatment

Treatment in this experiment consisted of two different kinds of tests: multiple-choice and short-answer. Using the textbook as a content base, the researchers, who were themselves all university professors knowledgeable in the area of educational psychology, first wrote an initial pool of multiple-choice test

items at two levels of complexity, (a) knowledge and (b) concept, following the guidelines given by Jenkins and Deno (1971). From this initial pool of items which bore no designation as to level, one of the researchers selected the 16 items which in her judgement were the best knowledge items and the 16 items which were the best concept items. The resulting 32 items were given to a second researcher who independently labled each item as being one which measured knowledge or concepts. The raters were in agreement on all items except two. Minor revisions were made in these two items to effect 100% interrater agreement on the knowledge-concept ratings for the items which were then randomly assigned their order of appearance on the test.

Each multiple-choice item had been written so that the stem alone would form a short-answer test item. The short-answer test for this experiment was thus composed of the stems of the multiple-choice test items with the alternatives omitted. Knowledge and concept levels of the short-answer items corresponded to the levels assigned to the respective multiple-choice items, and the random order of appearance which was used on the multiple-choice test was followed on the short-answer test.

Experimental Procedures

Subjects were randomly assigned to one of three groups, and each group was randomly designated as Treatment Group I--multiple-choice test; Treatment Group II--short-answer test; or Control Group--no test. On the day of the assigned classroom test on learning theory for which all participants had prepared, the instructor called the names of those assigned to the control group and asked that they meet briefly with her in another room. Here they were asked to participate in a study which involved their not taking the assigned test. For their participation in the study they would receive an

automatic grade of A on the test, and they were asked not to discuss the experiment with anyone until the instructor could explain it in class. All control group subjects agreed to these conditions and were subsequently dismissed.

The instructor returned to the classroom and routinely administered either a multiple-choice or a short-answer test to the respective treatment groups who were still naive with regard to the experiment. Both types of tests were dispensed from one common stack to further ensure the naivete of the experimental procedure. As each student completed his or her test, he or she turned it in to the instructor and left.

At the beginning of the next class day following the experimental test, all students were told by the instructor that their class had been selected to participate in a study of which their performance on the test they had just completed would be a part. They were also told that they would be given a detailed description of the procedure and results upon completion of the study, and for their participation they would be given a grade of A on that test.

Four weeks later, during which time the regularly scheduled classroom activities ensued, the instructor on an unannounced basis had all students respond to both experimental tests. The short-answer test was responded to prior to the distribution of and response to the multiple-choice test. Students were told that their performance on these tests was the concluding part of the experiment they had earlier been asked to participate in. They were encouraged to do their best on both tests with bonus points of a percentage of correct responses which would be added to their class grade.

Data Preparation and Analyses

Initial and delayed multiple-choice tests were scored for knowledge and for concepts using an objective scoring key. A detailed scoring key was prepared and utilized by the researchers to independently score the initial and delayed short-answer tests. Prior to the scoring all identifying data were concealed to guard against experimenter bias. Resulting interrater reliability co-efficients ranged from .98 to .99. Wherever there was disagreement on the scoring of a response, it was discussed in order to arrive at concensus.

Resulting scores were analyzed using a 3 x 2 x 2 factorial analysis of variance design with repeated measures in the last two factors. The between factor was treatment and the within factors were complexity level and type of question.

Results

Results of the analysis of variance are presented in Table 1. Means and standard deviations are in Table 2. As can be seen in Table 1, the treatment effect was not significant. Subjects in the multiple-choice test, short-answer test, and no test (control) groups did not differ in their delayed test performance (retention). Item level was also not significant. Retention scores did not differ on the conceptual and knowledge questions. Item type did result in a significant effect. Scores were higher on the multiple-choice test than on the short-answer test. The treatment conditions (taking different tests initially) did not interact with type of delayed test (multiple-choice vs. short answer) or with the complexity of the question (concept vs. knowledge). Further, the treatment by test type by question complexity interaction was nonsignificant. There was, however, an interesting and significant

finding in the interaction between test type and question complexity. Here the mean for conceptual questions was higher than the mean for rote memory questions on the multiple-choice test, but the mean for the conceptual questions was lower than the mean for rote memory questions on the short-answer test.

Conclusions

The results support the conclusion that the type of test used to measure student learning or retention does influence student performance as purported by Harnqvist (1977) and others. However, the degree of this influence may differ as a function of item complexity. Contrary to previous research (cf., Roderick and Anderson) the mere act of initially taking a test does not seem to significantly affect student performance at a later time (retention). It may be that it is the preparation for instead of the actual taking of the test that influences learning and retention. This speculation would need to be empirically tested, however.

Another interesting conclusion supported by this research is that conceptual learning does not seem to be retained any better than lower-level facts or knowledge. Such results are contrary to results of studies such as those of Tyler (1933) and Wert (1937) which suggested that concepts are remembered better than facts. A speculated reason to account for the present results would be that the students in this study overlearned both the factual and the conceptual material. Many studies dating back over 40 years (e.g., Krueger, 1929) have indicated that overlearning improves retention.

Educational or Scientific Importance of the Study

"To paraphrase a famous American, we have a dream--. . . . Our dream is of an educational system whose procedures are governed by research and by theories that are empirically based." Unfortunately this dream, representative of the desires of many but expressed by Dunkin and Biddle (1974) in their review of research on teaching, is far from being a reality at this time. For years educators have tried this teaching tactic, recommended that instructional approach and, too often, have had little or no evidence to back up these recommendations. How do we really know they work? The answer--we don't know in many cases!

Perhaps the beginning of wisdom is to discover how very little we know as yet if steps are taken to rectify that lack. The significance of this study lies in the fact that is one of the steps, however great or small, toward scientifically derived knowledge about the effects of classroom practices on student learning and retention.

Reference Note

1. Anderson, R. C., Surber, J. R., Biddle, W. B., Zych, P. M., & Lieberman, C. E. Retention of text information as a function of the nature, timing, and number of quizzes. Paper presented at the meeting of the American Educational Research Association, Chicago, April 1974.

References

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt Rinehart, and Winston, 1974.

Gall, M. D., Ward, B. A., Berliner, D. C., Cahen, L. S., Winne, P. H., Elashoff, J. D., & Stanton, G. C. Effects of questioning techniques and recitation on student learning. American Educational Research Journal, 1978, 15, 175-199.

Harnqvist, K. Enduring effects of schooling -- a neglected area in educational research. Educational Researcher, 1977, 6, 5 - 11.

Jenkins, J. R., & Deno, S. L. Assessing knowledge of concepts and principles. Journal of Educational Measurement, 1971, 8, 95-101.

Krueger, W. C. F. The effect of overlearning on retention. Journal of Experimental Psychology, 1929, 12, 71-78.

Peeck, J. Effect of prequestions on delayed retention of prose material. Journal of Educational Psychology, 1970 , 61, 241-246.

Roderick, M., & Anderson, R. C. A programmed introduction to psychology versus a text-book style summary of the same lesson. Journal of Educational Psychology, 1968, 59, 381-387.

Ryan, F. L. Differentiated effects of levels of questioning on student achievement. Journal of Experimental Education, 1973, 41, 63-67.

Tyler, R. W. Permanence of learning. Journal of Higher Education , 1933, 4, 203-204.

Wert, J. E. Twin examination assumptions. Journal of Higher Education, 1937, 8, 136-140.

Table 1

ANOVA of Treatment Group, Item Type, and Item Complexity

Source	df	Mean Square	F VALUE
Treatment	2	15.74	.64
Between Subjects Error	42	24.73	
Item type	1	1400.02	639.14***
Item type X Treatment	2	.24	.11
Within Subjects Error (1)	42	2.19	
Item Level	1	.02	.01
Item Level X Treatment	2	.61	.22
Within Subjects Error (2)	42	2.70	
Item Type X Item Level	1	13.89	7.24**
Treatment X Item type X Item Level	2	.51	.26
Within Subjects Error (3)	42	1.92	

** $\underline{P} < .01$
*** $\underline{P} < .001$

Table 2

Means and Standard Diviations by Treatment Group on Completion Knowledge, Completion Concept, Multiple-Choice Knowledge, and Multiple-Choice Concept Tests

Item Type/ Item Level	Group			
	Multiple-Choice	Completion	Control	Combined
Completion Knowledge				
Mean	7.47	6.80	6.20	6.82
Standard Deviation	3.48	3.30	2.43	3.08
Completion Concept				
Mean	6.53	6.60	5.73	6.29
Standard Deviation	2.80	3.38	2.96	3.01
Multiple-Choice Knowledge				
Mean	12.33	11.80	11.40	11.84
Standard Deviation	2.64	2.46	2.06	2.37
Multiple-Choice Concept				
Mean	12.93	12.47	11.87	12.42
Standard Deviation	2.76	2.59	2.47	2.59
Combined				
Mean	9.82	9.42	8.80	9.34